Papers

Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses

Fujian Song, Douglas G Altman, Anne-Marie Glenny, Jonathan J Deeks

Abstract

Objective To determine the validity of adjusted indirect comparisons by using data from published meta-analyses of randomised trials.
Design Direct comparison of different interventions in randomised trials and adjusted indirect comparison in which two interventions were compared through their relative effect versus a common comparator. The discrepancy between the direct and adjusted indirect comparison was measured by the difference between the two estimates.
Data sources Database of abstracts of reviews of effectiveness (1994-8), the Cochrane database of systematic reviews, Medline, and references of retrieved articles.
Results 44 published meta-analyses (from 28 systematic reviews) provided sufficient data. In most cases, results of adjusted indirect comparisons were not significantly different from those of direct comparisons. A significant discrepancy ($P < 0.05$) was observed in three of the 44 comparisons between the direct and the adjusted indirect estimates. There was a moderate agreement between the statistical conclusions from the direct and adjusted indirect comparisons (κ 0.51). The direction of discrepancy between the two estimates was inconsistent.
Conclusions Adjusted indirect comparisons usually but not always agree with the results of head to head randomised trials. When there is no or insufficient direct evidence from randomised trials, the adjusted indirect comparison may provide useful or supplementary information on the relative efficacy of competing interventions. The validity of the adjusted indirect comparisons depends on the internal validity and similarity of the included trials.

Introduction

For a given clinical indication, clinicians and people who make decisions about healthcare policy often have to choose between different active interventions. The number of active interventions is increasing rapidly because of advances in health technology. Therefore, there is an increasing need for research evidence about the relative effectiveness of competing interventions.

Well designed randomised controlled trials generally provide the most valid evidence of relative efficacy of competing interventions and minimise the possibility of selection bias.[1] However, many competing interventions have not been compared directly (head to head) in randomised trials. Even when different interventions have been directly compared in randomised trial(s) such direct evidence is often limited and insufficient.

As the results of placebo controlled trials are often sufficient to acquire the regulatory approval of new drugs, pharmaceutical companies may not be motivated to support trials that compare new drugs with existing active treatments. Lack of evidence from direct comparison between active interventions makes it difficult for clinicians to choose the most effective treatment for patients. In contrast, if clinicians and patients believe that the current treatment is effective, placebo controlled trials may be impossible, and new drugs are compared only with other active treatments in randomised trials. Therefore absence of direct comparison between new drugs and placebo may also be a problem.[2 3] For instance, the comparison of new drugs and placebo may be required for the purpose of regulatory approval and economic evaluations.

Because of the lack of direct evidence, indirect comparisons have been recommended[4] and used for evaluating the efficacy of alternative interventions (Glenny AM, et al, international society of technology assessment in health care, The Hague, 2000). There are concerns that indirect comparisons may be subject to greater bias than direct comparisons and may overestimate the efficacy of interventions.[5] Empirical evidence is required to assess the validity of indirect comparisons.

We previously examined the validity of indirect comparisons using examples in a systematic review of antimicrobial prophylaxis in colorectal surgery.[6] We found some discrepancies between the results of direct and indirect comparisons, depending on which indirect method was used. The results of the study, however, were based on only one topic and the findings may not be generalisable. We therefore used a sample of 44 comparisons of different interventions from 28 systematic reviews to provide stronger evidence about the validity of indirect comparisons.

Department of Public Health and Epidemiology, University of Birmingham, Birmingham B15 2TT
Fujian Song
senior research fellow

Centre for Statistics in Medicine, Institute of Health Sciences, Oxford OX3 7LF
Douglas G Altman
professor of statistics in medicine
Jonathan J Deeks
senior medical statistician

Cochrane Oral Health Group, University Dental Hospital of Manchester, Manchester M15 6FH
Anne-Marie Glenny
lecturer in evidence based oral health care

Correspondence to:
F Song
f.song@bham.ac.uk

P+

Details of methods and a worked example, references for 28 systematic reviews, and three tables are on bmj.com

Methods

To identify relevant meta-analyses of randomised controlled trials, we searched the database of abstracts of reviews of effectiveness (1994-8), the Cochrane database of systematic reviews (Issue 3, 2000), Medline, and references of retrieved articles. Our two inclusion criteria were that competing interventions could be compared both directly and indirectly and that the same trial data had not been used in both the direct and indirect comparison.

Comparison methods

The relative efficacy in each meta-analysis was measured by using mean difference for continuous data and log relative risk for binary data. We use two comparative methods: the direct (head to head) comparison and the adjusted indirect comparison. See webextra for more details about the statistical methods used and a worked example.

For the direct comparisons, comparison of the result of group B with the result of group C within a randomised controlled trial gave an estimate of the efficacy of intervention B versus C. We used the method suggested by Bucher et al for adjusted indirect comparisons.[5] Briefly, the indirect comparison of intervention B and C was adjusted by the results of their direct comparisons with a common intervention A. This adjusted method aims to overcome the potential problem of different prognostic characteristics between study participants among trials. It is valid if the relative efficacy of interventions is consistent across different trials.[4] [5]

When two or more trials compared the same interventions in either direct or indirect comparison, we weighted results of individual trials by the inverse of corresponding variances and then quantitatively combined them. According to our simulation studies, the adjusted indirect comparison using the fixed effect model tended to underestimate standard errors of pooled estimates (Altman DG, et al, third symposium on systematic reviews: beyond the basics, Oxford, 2000). Thus, we used the random effects model for the quantitative pooling in both the direct and the adjusted indirect comparison.[7]

Measures of discrepancy

The discrepancy between the direct estimate (T_{BC}) and the adjusted indirect estimate (T'_{BC}) was measured by the difference (Δ) between the two estimates:

$$\Delta = T_{BC} - T'_{BC}$$

the standard error being calculated by

$$SE(\Delta) = \sqrt{(SE(T_{BC})^2 + SE(T'_{BC})^2)}$$

where $SE(T_{BC})$ and $SE(T'_{BC})$ are the estimated standard errors for the direct estimate and the adjusted indirect estimate, respectively. The 95% confidence interval for the estimated discrepancy was calculated by $\Delta \pm 1.96 \times SE(\Delta)$. The estimated discrepancy can also be standardised by its standard error to obtain a value of $z = \Delta / SE(\Delta)$.

In addition, we categorised the results of meta-analyses as non-significant ($P > 0.05$) or significant ($P \leq 0.05$). The significant effect can be further separated according to whether intervention B was less or more effective than intervention C. The degree of agreement in statistical conclusions between the direct and indirect method was assessed by a weighted κ.[8]

Results

We identified 28 systematic reviews in which both the direct and indirect comparison of competing interventions could be conducted, although indirect comparison was not explicitly used in many of these meta-analyses. Some systematic reviews assessed more than two active interventions, and a total of 44 comparisons (see webextra table A) could be made by using data from the 28 systematic reviews.[w1-w28]

Figure 1 summarises the discrepancies between the direct and the adjusted indirect estimates. There was significant discrepancy ($P < 0.05$) in three of the 44 comparisons—that is, the 95% confidence interval did not include zero. In four other meta-analyses, the discrepancy between the direct and the adjusted indirect estimate was of borderline significance ($P < 0.1$). The relative efficacy of an intervention was equally likely to be overestimated or underestimated by the indirect comparison compared with the results of the direct comparison.

Figure 2 shows the relation between the statistical discrepancy (z) and the number of trials used in indirect comparisons. Visually, statistical discrepancies tended to be smaller when the number of trials was large (>60) than when the number of trials was small (<40). However, such a tendency was not consistent as the discrepancy between the direct and indirect estimate may be significant even when more than 40 trials have been used for the indirect comparison (see Zhang and Li Wan Po[w28]).

There was a moderate agreement in statistical conclusions between the direct and the adjusted indirect method (weighted κ 0.53) (table). In terms of statistical conclusions, 32 of the 44 indirect estimates fell within the same categories of the direct estimates. According to the direct comparisons, 19 of the 44 comparisons suggested a significant difference ($P < 0.05$) between competing interventions. Compared with direct estimates, the adjusted indirect estimates were less likely to be significant. Ten of the 19 significant direct estimates became non-significant in the adjusted indirect comparison, while only two of the 25 non-significant
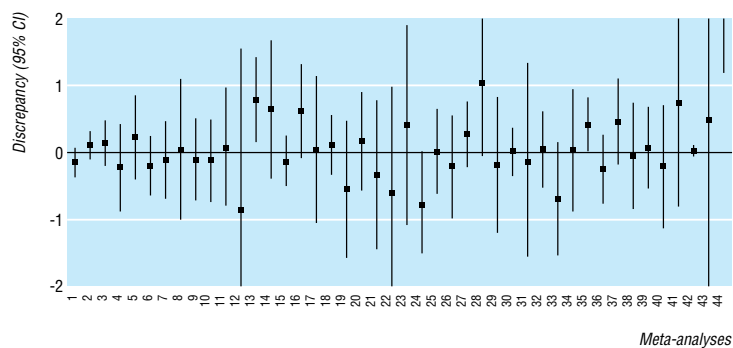
Fig 1 Discrepancy between direct and adjusted indirect comparison defined as difference in estimated log relative risk (meta-analyses 1-39) or difference in estimated standardised mean difference (meta-analysis 40) or difference in estimated mean difference (meta-analyses 41-44): empirical evidence from 44 published meta-analyses (see webextra table A)
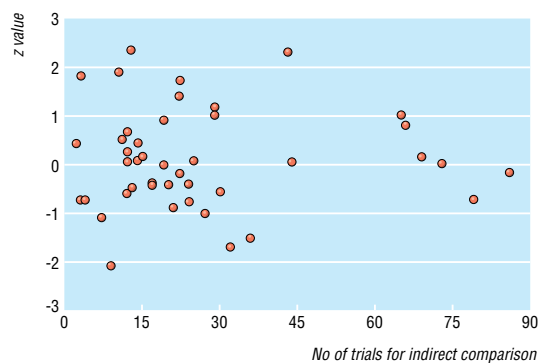
Fig 2 Statistical discrepancy (z value, calculated by dividing difference between direct and indirect estimates by its standard error ($z = \Delta / SE(\Delta)$)) and number of trials used in indirect comparison

Methods of comparison and number of significant findings* in 44 meta-analyses of competing interventions. Weighted κ 0.53 for agreement between direct and adjusted indirect estimate

	Adjusted indirect estimate		
Direct estimate	Significant effect (−) (n=6)	Non-significant effect (n=33)	Significant effect (+) (n=5)
Significant effect (−) (n=8)	5	3	0
Non-significant effect (n=25)	1	23	1
Significant effect (+) (n=11)	0	7	4

*Non-significant effect: difference between intervention groups is non-significant ($P > 0.05$); significant effect ($P \leqslant 0.05$) is separated according to whether intervention A is less (−) or more effective (+) than intervention B.

direct estimates was significant in the adjusted indirect comparison.

Discussion

The 44 meta-analyses in 28 systematic reviews included in this study covered a wide range of medical topics. The categories of patients included those with an increased risk of vascular occlusion, HIV infection, viral hepatitis C, gastro-oesophageal reflux disease, postoperative pain, heart failure, dyspepsia, and cigarette smoking. The results of adjusted indirect comparisons were usually similar to those of direct comparisons. There were a few significant discrepancies between the direct and the indirect estimates, although the direction of discrepancy was unpredictable. These findings are similar to (but more convincing than) those of our previous study of antibiotic prophylaxis in colorectal surgery.[6]

Discrepancies between the direct and the adjusted indirect estimate may be due to random errors. Partly because of the wide confidence interval provided by the adjusted indirect comparison, significant discrepancies between the direct and the adjusted indirect estimate were infrequent (3/44). Results that were significant when we used the direct comparison often became non-significant in the adjusted indirect comparison (table).

The internal validity of trials involved in the adjusted indirect comparison should be examined because biases in trials will inevitably affect the validity of the adjusted indirect comparison. In addition, for the adjusted indirect comparison to be valid, the key assumption is that the relative efficacy of an intervention is consistent in patients across different trials. That is, the estimated relative efficacy should be generalisable. Generalisability (external validity) of trial results is often questionable because of restricted inclusion criteria, exclusion of patients, and differences in the settings where trials were carried out.[9]

Of the 44 comparisons, three showed significant discrepancy ($P < 0.05$) between the direct and the adjusted indirect estimate. In two cases the discrepancies seem to have no clinical importance as both the direct and the adjusted indirect estimates were in the same direction.[w6 w17] However, the discrepancy between the direct and the adjusted indirect estimate was clini-

cally important in another case, which compared paracetamol plus codeine versus paracetamol alone in patients with pain after surgery.[w28] A close examination of this example showed that the discrepancy could be explained by different doses of paracetamol and codeine used in trials for the indirect comparison (box).
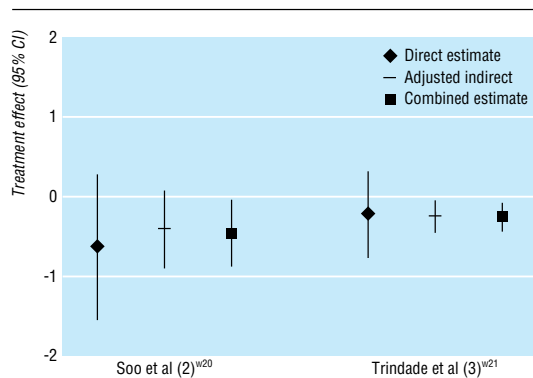
When is the adjusted indirect comparison useful?

When there is no direct evidence, the adjusted indirect method may be useful to estimate the relative efficacy of competing interventions. Empirical evidence presented here indicates that in most cases results of adjusted indirect comparisons are not significantly different from those of direct comparisons.

Direct evidence is often available but is insufficient. In such cases, the adjusted indirect comparison may provide supplementary information.[10] Sixteen of the 44 direct comparisons in this paper were based on one randomised trial while the adjusted indirect comparisons were based on a median of 19 trials (range 2-86). Such a large amount of data available for adjusted indirect comparisons could usefully strengthen conclusions based on direct comparisons, especially when there are concerns about the methodological quality of a single randomised trial.

Results of the direct and the adjusted indirect comparison could be quantitatively combined to increase statistical power or precision when there is no important discrepancy between the two estimates. The non-significant effect estimated by the direct comparison may become significant when the direct and the adjusted indirect estimate are combined, as happened

---

Importance of similarity between trials in adjusted indirect comparison

A meta-analysis by Zhang and Li Wan Po compared paracetamol plus codeine *v* paracetamol alone in postsurgical pain.[w28] Based on the results of 13 trials, the direct estimate indicated a significant difference in treatment effect (mean difference 6.97, 95% confidence interval 3.56 to 10.37). The adjusted indirect comparison that used a total of 43 placebo controlled trials suggested there was no difference between the interventions (−1.16, −6.95 to 4.64). The discrepancy between the direct and the adjusted indirect estimate was significant ($P = 0.02$). However, most of the trials (n=10) in the direct comparison used 600-650 mg paracetamol and 60 mg codeine daily, while many placebo controlled trials (n=29) used 300 mg paracetamol and 30 mg codeine daily. When the analysis included only trials that used 600-650 mg paracetamol and 60 mg codeine, the adjusted indirect estimate (5.72, −5.37 to 16.81) was no longer significantly different from the direct estimate (7.28, 3.69 to 10.87). Thus, the significant discrepancy between the direct and the indirect estimate based on all trials could be explained by the fact that many placebo controlled trials used lower doses of paracetamol (300 mg) and codeine (30 mg). This example shows that the similarity of trials involved in adjusted indirect comparisons should be carefully assessed.

Fig 3 Combination of direct and adjusted indirect estimates in two meta-analyses

What is already known on this topic

Many competing interventions have not been compared in randomised trials

Indirect comparison of competing interventions has been carried out in systematic reviews, often implicitly

Indirect comparison adjusted by a common control can partially take account of prognostic characteristics of patients in different trials

What this study adds

Results of adjusted indirect comparison usually, but not always, agree with those of head to head randomised trials

The validity of adjusted indirect comparisons depends on the internal validity and similarity of the trials involved

in two of the 44 comparisons (fig 3). In each case, the change was because of an increased amount of information.

It is also possible that the significant relative effect estimated by the direct comparison becomes non-significant after it is combined with the adjusted indirect estimate—for example, in a systematic review of H_2 receptor antagonists versus sucralfate for non-ulcer dyspepsia.[w20] The direct comparison based on one randomised trial found that H_2 receptor antagonist was less effective than sucralfate (relative risk 2.74, 95% confidence interval 1.25 to 6.02), while the adjusted indirect comparison (based on 10 trials) indicated that H_2 receptor antagonist was as effective as sucralfate (0.99, 0.47 to 2.08). The discrepancy between the estimates was marginally significant (P=0.07). The combination of the direct and adjusted indirect estimate provided a non-significant relative risk of 1.56 (0.93 to 2.75).

It will be a matter of judgment whether and how to take account of indirect evidence. It is not desirable to base such decisions on whether or not the difference between the two estimates is significant, although this is the easiest approach. A more constructive approach would be to base the decision on the similarity of the participants in the different trials and the comparability of the interventions.

Some authors have used a naive (unadjusted) indirect comparison, in which results of individual arms between different trials were compared as if they were from a single trial (Glenny AM, et al, international society of technology assessment in health care). Simulation studies and empirical evidence (not shown in this paper) indicate that the naive indirect comparison is liable to bias and produces overprecise estimates (Altman DG, et al, third symposium on systematic reviews: beyond the basics). The naive indirect comparison should be avoided whenever possible.

Direct estimates from randomised trials may not always be reliable

As has been observed, "randomisation is not sufficient for comparability."[11 12] The baseline comparability of patient groups may be compromised due to lack of concealment of allocation.[13] Patients may be excluded for various reasons after randomisation and such exclusions may not be balanced between groups. Lack of blinding of outcome measurement may overestimate the treatment effects.[13] Furthermore, empiri-

cal evidence has confirmed that published randomised trials may be a biased sample of all trials that have been conducted due to publication and related biases.[14]

Thus, direct evidence from randomised trials is generally regarded to be the best, but such evidence may sometimes be flawed. Observed discrepancies between the direct and the adjusted indirect comparison may be partly due to deficiencies in the trials making a direct comparison or those contributing to the adjusted indirect comparison, or both.

Conclusions

When there is no direct randomised evidence, the adjusted indirect method may provide useful information about relative efficacy of competing interventions. When direct randomised evidence is available but not sufficient, the direct and the adjusted indirect estimate could be combined to obtain a more precise estimate. The internal validity and similarity of all the trials involved should always be carefully examined to investigate potential causes of discrepancy between the direct and the adjusted indirect estimate. A discrepancy may be due to differences in patients, interventions, and other trial characteristics including the possibility of methodological flaws in some trials.

1 Pocock SJ. *Clinical trials: a practical approach*. New York: John Wiley, 1996.
2 Hasselblad V, Kong DF. Statistical methods for comparison to placebo in active-control trials. *Drug Inf J* 2001;35:435-49.

3 Fisher LD, Gent M, Buller HR. Active-control trials: how would a new agent compare with placebo? A method illustrated with clopidogrel, aspirin, and placebo. *Am Heart J* 2001;141:26-32.

4 McAlister F, Laupacis A, Wells G, Sackett D. Users' guides to the medical literature: XIX. Applying clinical trial results B. Guidelines for determining whether a drug is exerting (more than) a class effect. *JAMA* 1999;282:1371-7.

5 Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol.* 1997;50:683-91.

6 Song F, Glenny AM, Altman DG. Indirect comparison in evaluating relative efficacy illustrated by antimicrobial prophylaxis in colorectal surgery. *Contr Clin Trials* 2000;21:488-97.

7 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Contr Clin Trials* 1986;7:177.

8 Altman DG. *Practical statistics for medical research.* London: Chapman Hall, 1991.

9 Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;312:1215-8.

10 Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med* 1996;15:2733-49.

11 Abel U, Koch A. The role of randomization in clinical studies: myths and beliefs. *J Clin Epidemiol* 1999;52:487-97.

12 Cappelleri JC, Ioannidis JPA, Schmid CH, Ferranti SDd, Aubert M, Chalmers TC, et al. Large trials vs meta-analysis of smaller trials. How do their results compare? *JAMA* 1996;276:1332-8.

13 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.

14 Song F, Eastwood A, Gilbody S, Duley L, Sutton A. Publication and related biases. *Health Technol Assess* 2000;4:1-115.